COMPARING STUDENT PERFORMANCE ON COMPUTER-BASED VS. PAPER-BASED TESTS IN A FIRST-YEAR ENGINEERING COURSE

Meagan E. Ita, Krista M. Kecskemety, Katlyn E. Ashley, Brooke C. Morin
Engineering Education Innovation Center
The Ohio State University

Abstract

With the increasing prevalence of computer-based examinations in universities, the Fundamentals of Engineering for Honors (FEH) program at The Ohio State University conducted a study examining the effect of test mode on student exam performance. Computer-based exams have many advantages for the instructor over paper-based exams including automated grading, ease of distribution, reduced paper usage, and the ability to see student exam statistics instantly. This study was conducted during the fall semester and consisted of two midterm exams and a final exam. The exam questions being considered in this study included multiple choice, true/false, multiple select, and fill in the blank questions. For each exam, there were two parts (1 and 2), that each student took with separate modes, either on computer or on paper. The part (1 or 2) each student took in each mode was flipped for half the students. After analyzing the results, the exam score differences between computer and paper were not statistically significant in 5 out of 6 exam parts analyzed. While the exam scores displayed limited differences, the students reported an overwhelming preference for the paper exams (86%) over the computer exams (5%). These results indicate that even though computer exams have similar performance to their paper counterparts, the students prefer the paper exams. Therefore, because performance is similar, future studies should explore improving the student experience in the computer exam environment while still realizing the advantages of computer-based examinations.

Introduction

Computer-based examinations are commonly utilized in the university setting as well as in other areas, such as government-related examinations and standardized testing. Additionally, computer-based exams are increasingly common in the university setting along with the increasing prevalence of online textbooks, course content databases, and homework/other non-timed online assessments. Computer-based testing is an attractive option for educators for many reasons, including test distribution logistics and mitigation of time and work requirements, especially when evaluating large groups of students. In order to validate computer-based testing over more traditional paper-and-pencil testing, however, it is important to ensure that difference in test mode does not have an adverse effect on student performance.

While computer-based testing can be beneficial from an educator perspective, the literature regarding this testing mode from a student performance perspective is controversial. Studies that analyze test mode effect (effects of computer versus paper-based test mediums) are conflicting and do not always find a favorable result for computer-based tests. For example, in 1989 Bunderson, et al. highlighted 23 studies that evaluated test mode effect and reported 9 studies that favored paper-based, 3 studies that favored computer-based, and 11 studies that reported no difference between test modes [1]. Regardless of the advance of technology in the 25 years since this publication, the comprehensive conclusions of test mode effect studies remain debatable. Lee and Weerakoon [2] (2001) and Russel [3] (1999) both reported enhanced student

performance on paper over computer while Clariana and Wallace [4] (2002) reported enhanced performance on computer over paper. Still others continue to report no difference[5,6].

In considering the inconsistent findings in the literature, one must also consider the variable research designs and sample populations used within these studies. For example, sample populations have ranged from 137 college level students evaluated in chemistry [6] to over 1,100 students of various educational backgrounds evaluated by the Graduate Record Examination (GRE) [7]. Both studies attempted to relate gender to test mode. While the former found no correlation, the later sample population did find a correlation between gender and test mode. This example portrays how test mode effects can vary based on sample population both in terms of number and subjects. Additionally, research designs vary across test mode effect studies in respect to what variables are controlled. For example, Ricketts and Wilks conducted back to back studies on student populations where, in the first, aesthetics between computer and paper tests were not tightly controlled, but in the second study measures were taken to more closely match aesthetics [8]. Specifically, these researchers found that student performance improved in an online assessment when they were not required to scroll through the online test. The many variables involved in these test mode effect studies articulate that educators must carefully consider the relative circumstances of their student populations when applying results from these test mode effect studies, as results from a certain study may not appropriately translate. This same caution is stressed as educators interpret the results of the study presented in this paper.

Some studies have also correlated certain qualitative data to test mode effect and articulated confounding variables of concern. Clariana and Wallace, who found that computer-based test delivery positively impacted test scores of college students in a Computer Fundamentals course, analyzed gender, computer familiarity, competitiveness, and content familiarity relative to test mode [4]. They found that of these factors, only content familiarity was associated with test mode effect. Additionally, results showed that high-attaining students gained a larger advantage with computer-based relative to paper-based testing. A study conducted with first year engineering students in a chemistry course in Turkey reported no significant difference between paper- and computer-based test modes [5]. This study highlighted three confounding variables of concern in test mode effect studies including personal characteristics of test takers, various features of computer-based testing systems, and test content. These researchers believed that once these various factors are controlled, test mode effect can be eliminated.

To our knowledge, there have been no test mode studies conducted with engineering students *in an engineering course*. Additionally, the rapid advance of technology and incorporation into students' lives at earlier ages certainly plays a role in how students may approach a paper-based versus a computer-based test. With this in mind, it is important to gather up-to-date data on students with the described demographic. We believe that analyzing test mode effect with first-year engineering students in an engineering course could therefore contribute to many university-level engineering programs. The findings of this study will also benefit the students in the FEH program at The Ohio State University. By analyzing test mode effect in routine midterm and final examinations and isolating confounding factors that correlate to test mode effect, we can structure exams in order to maximize student learning, retention, and success.

Therefore, an honors first-year engineering program at The Ohio State University conducted a study to examine the test mode effect for two midterm exams and one final exam. This study sought to address the following research question: *Is there a difference in student performance between computer and paper*

based exams, and, if so, what factors contribute to any differences?

Methods

Description of Course

This study was conducted with students enrolled in the first of two courses of the FEH program. In this first-semester course, students studied problem solving by utilizing algorithm development and computational tools such as Excel, MATLAB, and C++ programming. The course also covered the topics of academic integrity, engineering ethics, data analysis, team building, and the engineering design process. The grade breakdown for the course is shown in Table 1.

Table 1: Grade breakdown for a first-year engineering, first semester course.

Grade Category	**% of Grade**
Preparation Assignments	10%
Application Assignments	20%
Laboratory Assignments	21%
Journals	3%
Design Project	5%
Extra Weekly Assignments	3% BONUS
Quizzes	6%
Midterm Exams	20%
Final Exam	15%

There were 13 total class sections of this course made up by 11 standard sections, 1 advanced programming section, and 1 section that utilized LabVIEW as a programming language instead of the C++ programming language. Because course content for the LabVIEW exam deviated from the other sections after the first midterm examination, data from students in the LabVIEW section were only included in the Midterm 1 results.

Design of the Study

Approximately 360 students participated in this study. Data were collected from a total of three examinations: two midterm exams and one final exam. The portion of each exam from which grades were collected for data were split into two parts, Part 1 and Part 2, which together comprised 40-50% of each total exam grade. The question types for this portion of the exam included multiple-choice, multiple-select, true or false, and fill in the blank. In each class, half of the students completed Part 1 on paper and Part 2 on the computer, while the other half of students did the reverse. Part 3 of each examination can be considered as the remaining 50-60% of each test and was completed in the same mode for both versions (A and B). Most of Part 3 was completed on paper; however, some questions required the use of a computer. The results from Part 3 were not considered in this test mode study. The test breakdown is represented schematically in Figure 1, and represents the same structure of all three examinations used in this study.

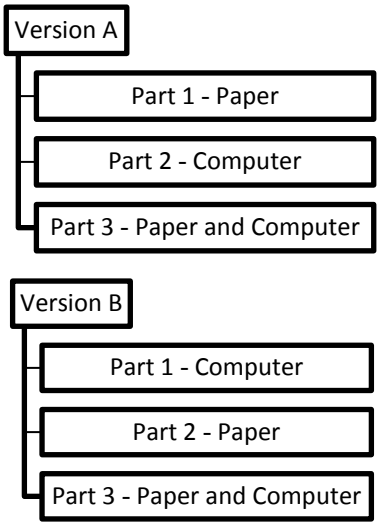


Figure 1: Schematic of test breakdown.

Question phrasing and order on the same part (1 or 2) were identical, regardless of testing mode. The paper portions were created to be as aesthetically similar as possible to the computer portions, which were completed on a course management system. This course management system, Carmen, is a university specific system developed by Desire2Learn. The aesthetic similarities included using the same font and making the figures appear the same in both test modes. Students were familiar with completing

untimed quizzes in this environment prior to the exams; however, before the first midterm examination they had never experienced the quiz environment in a timed manner. On the computer portion of the exam, students were able to freely navigate between questions to allow the same "flip-back" opportunity available with paper-based questions. This allowed students to change an answer on the computer portion of the exam, similar to how students would be able to change an answer on the paper portion. In order to minimize scrolling, the students were advised to maximize their browser window, although this was not required. Additionally, the computer portion had fewer questions per page in order to typically not require scrolling when the window was maximized.

To determine the test mode preference of the students, an end-of-semester course survey included the following question:

Q. Throughout this semester, you had exams in [course number] that had multiple choice, true/false, and fill-in questions both on paper and online. Which of these methods do you prefer for exams?

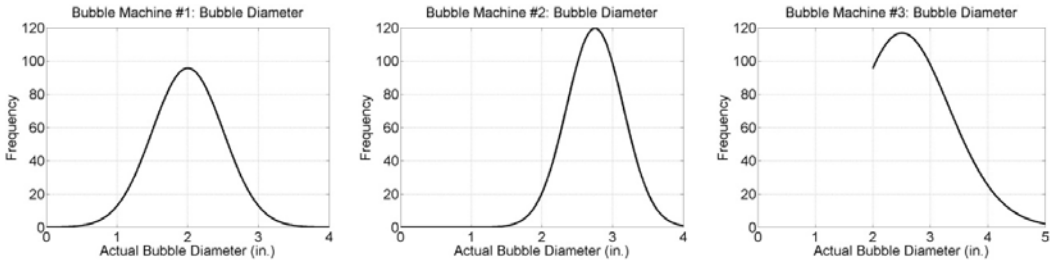
- *On Paper*
- *Online*
- *No Preference*

Note, the question was phrased with "Online" as a choice instead of "Computer" in order to differentiate the questions taken in Part 1 or 2, which are part of this study, and any questions they took in Part 3.

Sample Questions

When designing the exam there were a few options in the format of the paper portion. One option was to take screen shots of the computer questions and use print-outs of these screenshots for the paper portion of the exams. While this would have made the exams aesthetically similar, it would not have represented the normal process for an instructor writing an exam, would have used more paper, and may have been more confusing for students to select their answers. Therefore, the questions were developed in Microsoft Word rather than using a screen shot from the learning management system. Examples of the same two true/false questions from the paper and computer portion of the exams are shown in Figure 2 and Figure 3, respectively.

(Q1-Q5) Three bubble producing machines intended to produce uniform bubbles with 2 in. diameters have been designed. The frequency distribution plots of the actual measured diameters are below. Use these plots to answer the next five questions.



1. (2pts) True or False: Machine 1 is more precise than Machine 2. **Answer:** _____

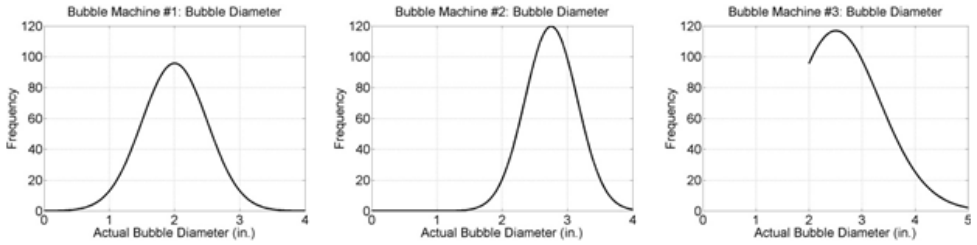
2. (2pts) True or False: Machine 1 is more accurate than Machine 2. **Answer:** _____

Figure 2: Sample true/false questions - paper portion of examination.

Information

(Q1-Q5) Three bubble producing machines intended to produce uniform bubbles with 2 in. diameters have been designed. The frequency distribution plots of the actual measured diameters are below. Use these plots to answer the next five questions.

Information



Question 1 (2 points)

True or False: Machine 1 is more precise than Machine 2.

- True
- False

Save

Question 2 (2 points)

True or False: Machine 1 is more accurate than Machine 2.

- True
- False

Save

Figure 3**:** Sample true/false question - computer portion of examination.

Statistical Analysis

Overall results were compared in each test mode for each exam. Since the results were typically non-normal distributions, a Wilcoxon rank-sum test was used to compare the test mode results.

In addition to examining the overall results, individual student results were examined over the three exams across testing modes. To limit the scope of this investigation, only two course sections were used. The two course sections chosen had different instructors; therefore, the instructor effect could be investigated. Additionally, these course sections were chosen because they did not have the highest or lowest exam averages of the sections. The student results were normalized to the course section averages rather than the overall course averages to eliminate any impact from the "course section" choice.

In order to investigate individual student performance consistency, each student was given a "Paper Performance Score" which indicated how much better he or she performed on the paper portion compared to the computer portion of the exam. The equations used to calculate this "Paper Performance Score" (PPS) are shown in Equations 1 and 2. An example of this calculation can be found in Appendix A.

Student X took Version A Exam (Part 1 (P1) on paper and Part 2 (P2) on the computer)
Student Y took Version B Exam (Part 2 (P2) on paper and Part 1 (P1) on the computer)

$$Student\ X\ PPS = \frac{(P1_{score} - P1_{class\ average})}{P1_{class\ average}} - \frac{(P2_{score} - P2_{class\ average})}{P2_{class\ average}} \quad (1)$$

$$Student\ Y\ PPS = \frac{(P2_{score} - P2_{class\ average})}{P2_{class\ average}} - \frac{(P1_{score} - P1_{class\ average})}{P1_{class\ average}} \quad (2)$$

Because the course section average of each part was used for normalization, this minimized any effect of different content appearing in Part 1 and Part 2. This measure enabled the testing of the consistency to which students had a more favorable test mode environment. A student who has a positive score for all 3 exams would show a more favorable result with paper exams, where as a student with a negative score for all 3 exams would demonstrate a more favorable result with computer exams.

Results and Discussion

Overall student score averages on both paper and computer portions of Part 1 and Part 2 for Midterm 1, Midterm 2, and the Final Exam are shown in Table 2. A Wilcoxon rank-sum test was used to test for a significant difference between the paper and computer versions for each of the 2 parts for all three examinations, for a total of six tests for statistical significance. The resulting p-values are shown in Table 2. A significant difference was found between the paper and computer portions of Part 2 on Midterm 2. This difference favored the paper version, as shown by the shaded cell in Table 2

To evaluate student performance between paper and computer versions with respect to student testing mode preference, the sample population of students was divided into those students who designated a paper testing preference and those who designated a computer testing preference. Overall score averages were calculated for each subset population, as well as p-values for a Wilcoxon rank-sum test for a significant difference between testing modes. Results for those students who preferred paper and for those who preferred computer (worded as "online" in the student survey) are shown in Table 3 and Table 4, respectively. Students who did not indicate a preference were not included in these subset populations.

Table 2: Overall results (N=370, N=337, N=332), p-value computed using a Wilcoxon rank-sum test. A significant difference is represented by a p-value < .05 and designated by **.

	Part 1			Part 2		
	Paper Average	**Computer Average**	**p-value**	**Paper Average**	**Computer Average**	**p-value**
Midterm 1	88.26%	87.75%	0.7572	79.94%	79.81%	0.8709
Midterm 2	69.85%	69.62%	0.8224	74.03%	70.62%	0.0417**
Final Exam	86.48%	86.19%	0.5381	80.73%	78.94%	0.2136

Table 3: Results for students who preferred paper (N=302, N=277, N=277), p-value computed using a Wilcoxon rank-sum test. A significant difference is represented by a p-value < .05 and designated by **.

	Part 1			Part 2		
	Paper Average	**Computer Average**	**p-value**	**Paper Average**	**Computer Average**	**p-value**
Midterm 1	88.13%	87.53%	0.7475	80.51%	79.22%	0.3243
Midterm 2	69.63%	69.93%	0.9416	73.66%	70.18%	0.0410**
Final Exam	86.10%	85.92%	0.7204	80.86%	78.01%	0.0586

Table 4: Results for students who preferred computer (N=16, N=14, N=14), p-value computed using a Wilcoxon rank-sum test. A significant difference is represented by a p-value < .05 and designated by **.

	Part 1			Part 2		
	Paper Average	**Computer Average**	**p-value**	**Paper Average**	**Computer Average**	**p-value**
Midterm 1	85.83%	91.25%	0.2679	74.38%	84.38%	0.0300**
Midterm 2	65.56%	64.00%	0.9161	73.17%	66.94%	0.6993
Final Exam	89.81%	86.67%	0.5455	71.67%	81.01%	0.1129

For the students who preferred the paper testing mode, a significant difference was found between the paper and computer portions for the same test condition that showed a significant difference for the total population: Part 2 on Midterm 2. This makes sense because the students who preferred paper made up a majority of the total student sample population. Only one test condition tested positive for statistical significance for the subset of students who preferred computer as well, although the portion exhibiting a difference was of Part 2 on Midterm 1, rather than Midterm 2. It should be noted that the subset of students who preferred paper is around 10 times as large as the subset of students who preferred computer, as indicated by Figure 4. The students who preferred computer resulted in a small sample size for the statistics calculated in Table 4, and therefore results with the computer preference population may be less reliable than those for the paper preference population.

While significant differences were found for one test condition for each the total student population and each subset population of exam preference (Tables 2 - 4), the majority of the results show no statistically significant difference between paper and computer versions. Additionally, although almost 90% of students indicated that they preferred paper-based examinations over computer-based examinations, the student performance results do not reflect this test mode preference.

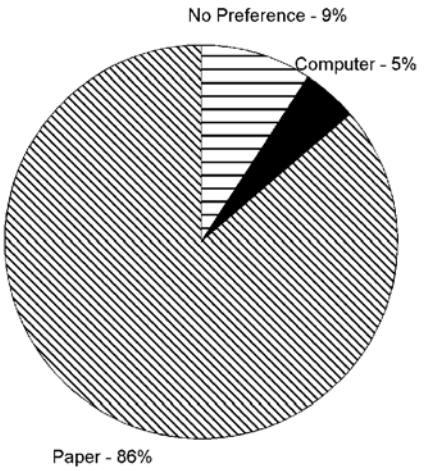


Figure 4: Breakdown of student testing mode preference (N = 351).

Interestingly, in a study conducted by Koch and Patience where Likert-type scales regarding general test preference were administered to students, students preferred computer-based tests more often than paper-based [9]. Similar

to results shown here, no correlation could be made between student test mode preference and exam performance.

Individual exam questions were additionally analyzed for statistical significance using a Wilcoxon rank-sum test. From all three exams, five questions exhibited significantly different student performance between computer and paper portions of identical questions. These five questions, a description of each, computer and paper averages, and their p-values are shown in Table 5. The test mode in which students showed higher performance (paper or computer) is shaded.

Of the individual exam questions outlined in Table 5, four of the five of these questions involved a figure or answer bank that the student had to refer back to in order to answer the question. While we expected questions which referred back to figures, etc. to favor performance in the paper version of the exam, the results for these four questions are exactly split, with two of the questions scoring significantly higher on paper and two scoring significantly higher on the computer. Further, the two questions from Part 2, Midterm 2 referred to the same answer bank and were of similar format, but resulted in higher student performance with opposite testing modes.

To evaluate performance consistency between different testing modes of individual students, the Paper Performance Score (PPS), as described in Methods, was calculated for two course sections. PPS was plotted per student for Course Section 1 and 2 as shown in Figure 5 and Figure 6, respectively, to explore whether individual students consistently performed better in one format over the other. A positive score indicates better performance on the paper version over the computer version of the respective exam, while a negative score indicates better performance on the computer version. Bar plots for each course are divided by solid vertical lines according to student performance across all three exams. For example, the 1st region (far left) includes students who consistently scored higher on paper versions and the 4th region (far right) includes students who consistently scored higher on the computer versions.

Figure 5 indicates that three students consistently scored higher on paper and three students consistently scored higher on the computer in Course Section 1, while Figure 6 indicates that two students consistently scored higher on paper and three students consistently scored higher on computer in Course Section 2.

Table 5: Questions which demonstrated statistically significant differences ($p < .05$) computed using a Wilcoxon rank-sum test.

Question	Description	Paper Average	Computer Average	p-value
Midterm 1, Part 1 Question 3	T/F, Given a distribution plot, determine if the results show systematic error.	82.87%	92.74%	0.0043
Midterm 2, Part 2 Question 4	MC, Given a bank of Unix/C commands match to description.	88.89%	79.63%	0.0202
Midterm 2, Part 2 Question 9	MC, Given a bank of Unix/C commands match to description.	87.72%	94.44%	0.0325
Final, Part 1 Question 19	MC, Given two histograms, which has the greater mean.	93.29%	85.71%	0.0249
Final, Part 2 Question 18	Fill in the Blank, give the Excel command to type in a cell.	88.39%	82.32%	0.0051

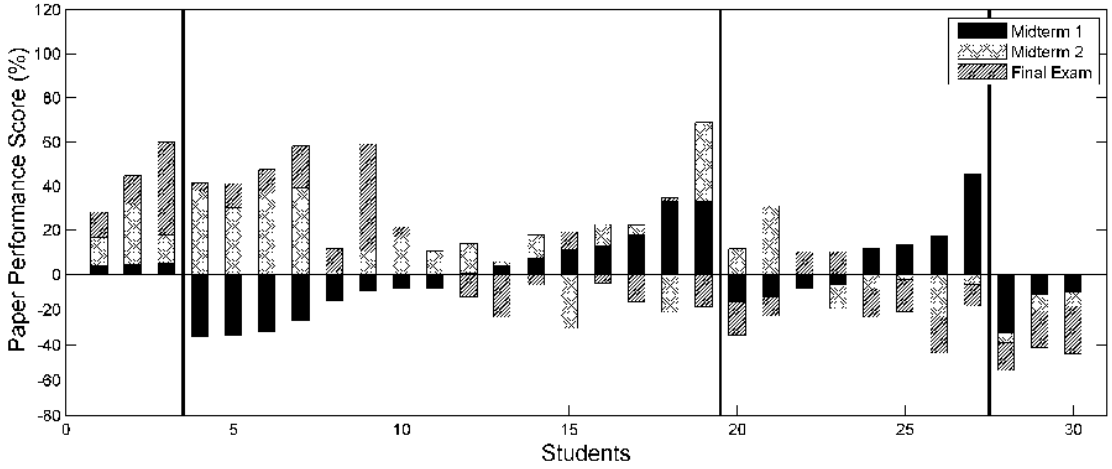


Figure 5: Course Section #1. The four regions of this plot indicate in order from left to right: 3 exams favoring paper, 2 exams favoring paper and 1 exam favoring the computer, 1 exam favoring paper and 2 exams favoring the computer, and 3 exams favoring the computer.

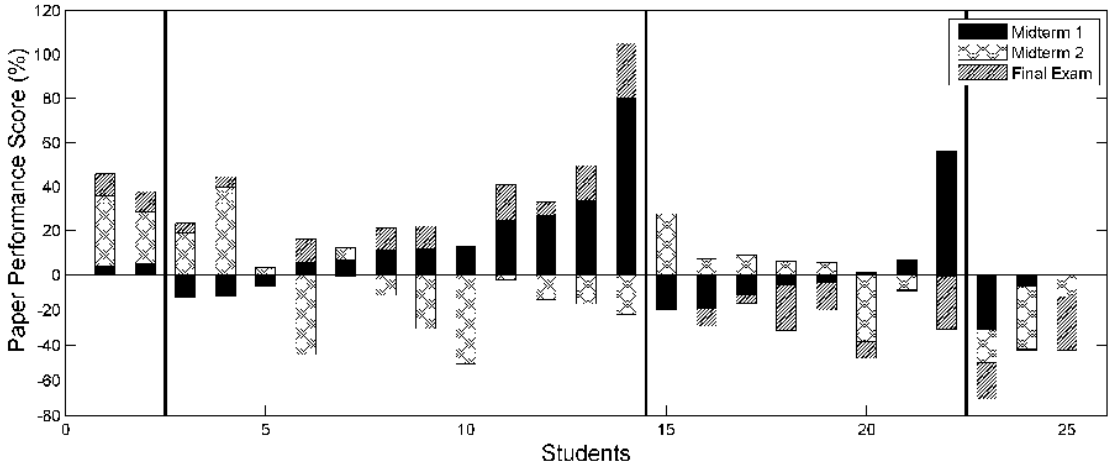


Figure 6: Course Section #2. The four regions of this plot indicate in order from left to right: 3 exams favoring paper, 2 exams favoring paper and 1 exam favoring the computer, 1 exam favoring paper and 2 exams favoring the computer, and 3 exams favoring the computer.

Overall results of these two plots, however, do not indicate any clear trend of consistent student performance with either test mode. Additionally, when comparing the course sections, there are similar proportions of students in the four regions.

Conclusions

A test mode effect study was conducted with first-year engineering students in a Fundamentals of Engineering for Honors course at The Ohio State University. Student score data were analyzed for both paper and computer portions of two midterm exams and one final exam to determine if test mode affected student performance. Although a statistically significant difference was found between paper and computer portions of one part of one midterm exam, overall results show no difference between testing mode. Subsets of the sample population also exhibited overall randomness and no clear favoring of one test mode over another. Variable results within the

data set analyzed here are similar with the literature previously outlined; if we consider test mode effect studies collectively, it is hard to conclude definitively whether test mode affects student test performance.

While student preference was not reflected in student performance for this study, student preference for test mode and its effect on test performance should be further studied. We feel it is important to further investigate reasons behind students' high preference toward paper-based examinations in order to understand why the preference toward paper exists. Specifically, we would like to isolate what exactly about paper-based exams students prefer and why. While we did not collect opinions from students regarding why they preferred one testing mode over the other, we can speculate student reasoning based on the experimental methods. For example, efforts were made to make the paper-based and computer-based exams as aesthetically identical as possible. If we therefore consider aesthetic differences as negligible, then we can focus on functional differences between the test modes that could contribute to preference of one or the other. For example, students may find changing pages easier in either the paper or computer test mode, or may prefer the testing technique of marking directly on a paper test. One step toward isolating the reason behind student preference could be to retroactively analyze completed tests to look for markings students made on paper exams during testing. High student preference toward paper exams could translate to students having a more comfortable testing experience. Future work in this area could include giving students a paper exam, while requiring students to enter exam answers online. This testing technique could accommodate student preference while still providing advantages of electronic grading, electronic statistics, and feedback to the instructors.

Because the survey regarding student preference did not take place until the final course evaluation, we were unaware of the students' overwhelming preference toward paper exams throughout the semester and were unable to factor that concern into the study. With regards to addressing this preference in the future, the debate is still ongoing in the course as to the format for future examinations. Course instructors are leaning toward continuing the computer testing for the ease of grading, and since the results showed similar scores for computer and paper testing, it is believed that continuing to test students on the computer will not adversely affect students' grades. The instructors are, however, exploring possible ways to positively impact the student experience during computer exams to make students more comfortable with the computer-based testing environment. Any changes will be influenced by a future study investigating why students overwhelmingly prefer paper-based exams.

While we have shown that there is no significant difference between test mode and student performance in all but one of the test score sets analyzed, the sample population and course content should be considered with this conclusion. As with previously conducted test mode effect studies, results presented here may not be applicable to all populations and courses and should be looked at subjectively. The content of the course involved in this study was encompassed by mainly computer programming applications. Future test mode effect studies should be conducted in other engineering courses spanning different engineering disciplines and including more widespread content.

Acknowledgements

The authors would like to acknowledge the Fundamentals of Engineering for Honors instructional staff for their assistance in administering this research study. Also, the authors would like to thank Mr. Paul Clingan, Dr. Deb Grzybowski, and Dr. Kathy Harper for their advice in the initial stages of this study. Additionally, the authors would like to thank Samantha Nehrbass for her help collecting and processing the data.

References

1. C.V. Bunderson, D.K. Inouye, and J.B. Olsen, "The four generations of computerized educational measurement," *Educational measurement: American Council on Education*, Washington DC, 1989.

2. G. Lee and P. Weerakoon, "The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests," *Medical Teacher*, vol. 23, no. 2, 2001.

3. M. Russell, "Testing on computers: A follow-up study comparing performance on computer and on paper," *Education Policy Analysis Archives*, vol. 7, no. 20, 1999.

4. R. Clariana and P. Wallace, "Paper-based versus computer-based assessment: key factor associates with the test mode effect," *British Journal of Educational Technology*, vol. 33, no. 5, 2002.

5. S. Ozalp-Yaman and N.E. Cagiltay, (2010). "Paper-Based versus Computer-Based Testing in Engineering Education," *IEEE EDUCON Education Engineering 2010*, 1631-1637, 2010.

6. H.S. Ashton, C.E. Beevers, A.A. Korabinski, and M.A. Youngson, "Investigating the medium effect in computer-aided assessment of school Chemistry and college Computing national examinations," *British Journal of Educational Technology*, vol. 36, no. 5, 2005.

7. C. Parshall and J.D. Kromery, "Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect.," *Annual Meeting of the American Educational Research Association*, Atlanta, GA. 1993.

8. C. Ricketts and S.J. Wilks, "Improving Student Performance Through Computer-based Assessment: insights from recent research," *Assessment and Evaluation in Higher Education*, vol. 27, no. 5, 2002.

9. B.R. Koch and W.M. Patience, "Student attitudes toward tailored testing," *Proceedings of the 1977 Computerized Adaptive Testing Conference*. Minneapolis, MN, 1978.

Biographical Information

Meagan Ita attended The Ohio State University where she received a BS and MS in Biomedical Engineering, with a research focus in injury biomechanics. Meagan was a teaching assistant with the Engineering Education Innovation Center (EEIC) at The Ohio State University for the Fundamentals of Engineering with Honors program from 2010 to 2014. She is interested in investigating first-year experiences in engineering and optimizing the learning experience for these students. She will attend the University of Pennsylvania in the fall of 2014 to pursue a PhD in Bioengineering.

Krista Kecskemety is a senior lecturer in the Engineering Education Innovation Center at The Ohio State University. Krista received her B.S. in Aerospace Engineering at The Ohio State University in 2006 and received her M.S. from Ohio State in 2007. In 2012, Krista completed her Ph.D. in Aerospace Engineering at Ohio State. Her engineering education research interests include investigating first-year engineering student experiences, faculty experiences, and the connection between the two.

Katlyn Ashley is a student currently pursuing a B.S. in Chemical Engineering at The Ohio State University. Katlyn is also an Undergraduate

Teaching Assistant in the Engineering Education Innovation Center at Ohio State, which prompted her research interests in first-year engineering education.

Brooke Morin has worked as a Lecturer in the College of Engineering at The Ohio State University, teaching First-Year Engineering for Honors classes in the Engineering Education Innovation Center. Brooke earned her bachelor's degree and master's degree in Mechanical Engineering at Ohio State. Her research interests include first-year engineering education best practices and teaching assistant training.

Appendix A: Sample Paper Performance Score Calculation

To demonstrate how to calculate the Paper Performance Score, the following method was used:

Student X took Version A Exam (Part 1 (P1) on paper and Part 2 (P2) on the computer)
Student Y took Version B Exam (Part 2 (P2) on paper and Part 1 (P1) on the computer)

To calculate the "Paper Performance Score" (PPS), Equations 1 and 2 were used:

$$Student\ X\ PPS = \frac{(P1_{score} - P1_{class\ average})}{P1_{class\ average}} - \frac{(P2_{score} - P2_{class\ average})}{P2_{class\ average}} \quad (1)$$

$$Student\ Y\ PPS = \frac{(P2_{score} - P2_{class\ average})}{P2_{class\ average}} - \frac{(P1_{score} - P1_{class\ average})}{P1_{class\ average}} \quad (2)$$

Consider the given scores for Students X and Y in Table 6 and assume Part 1 and Part 2 class averages were 79 and 76, respectively. Given this information, the calculated PPS for Students X and Y is shown below.

Table 6: Sample calculation of "Paper Performance Score".

	$P1_{score}$	$P2_{score}$	$P1_{score}$ % $diff$	$P2_{score}$ % $diff$	PPS
Student X	78	82	(78-79)/79 = –1%	(82-76)/76 = +8%	– 9%
Student Y	67	66	(67-79)/79 = –15%	(66-76)/76 = –13%	+ 2%

From this example it would appear that Student X performed better on the computer portion, and Student Y had a slightly more favorable outcome with the paper portion.